AUTOMATED HEALTH CARE MANAGEMENT SYSTEM USING BIG DATA TECHNOLOGY

Ms.A.Sivasankari.,Mrs.N.Sindhuja.,Mrs.S.Selvakani.,

[1]*Head of the department, Department of Computer science, DKM College for Women (Autonomous), Vellore.*

[2]*Research scholar,Department of Computer Science, DKM College for Women (Autonomous), Vellore, TamilNadu.*

[3]*Department of Computer Science,Thiruvalluvar university college arts and science,Arakkonam,Tamilnadu.*

---***---

ABSTRACT – Generally Automation plays an important role in the global economy and in daily experience. The Automated Healthcare Management System is an automated system that is used to manage patient information and its administration. In existing system challenges at large scale performing large-scale computation is difficult. To work with this volume of data requires distributing parts of the problem to multiple machines to handle in parallel. The information gained from analyzing massive amounts of aggregated health data can provide useful insight to improve quality and efficiency for providers and insurers a like.This makes the patients reach out for healthcare solutions easily and cheaply and makes healthcare a easy reach for the unprivileged also. Thus, this unified model can serve as a data collection, delivery as well as an analytic tool in the healthcare domain.This paper addresses the problem of data quality in electronic patient records using a computerized patient records report system with Apache HIVE and abstraction of Map reduce of big data technology. We analyzed which patient is spending more money than the others with the Map reduce. We got the data to be processed from traditional system to Hadoop via ETL's. We organized this with Oozie scheduler in Hadoop. The data what you are going to analyze is an semi-structured data. After uploading their data to cluster anyone can access them again provided they got to be in the cluster or can also use virtual machines that contain the right software to analyze them without any need for conversion.

Key words – AHMS, Oozie, HIVE, Mapreduce,ETLs,Hadoop.

1.INTRODUCTION

Generally Automation plays an important role in the global economy and in daily experience. Engineers trying to combine automated devices with mathematical and organizational tools to create complex systems for a rapidly expanding range of applications. The Automated Healthcare Management System is an automated system that is helps to manage patient information and its administration. It is meant to give the Administration and Staff, with information in practical to make their work more interesting and less stressing.

Whenever multiple machines are used in cooperation with one another, the probability of failures rises. In a single-machine environment, failure is not something that program designers explicitly worry about very often: if the machine has crashed, then there is no way for the program to recover anyway. This is because of the fact that a set of particular symptoms will not always lead to a particular disease and can be causing another set of diseases.

So we may be able to tap the appropriate set of diseases linked to the symptoms easily after analysis. This model also meets the key challenges posed to us in health sector that are shortage of human resources in the sector, accessibility of healthcare infrastructure, affordability of healthcare services especially for the rural population. Some major benefits from the model includes:

•Cutting down recurring medical costs

•Well- maintained medical history

Secured medical records accessible any-time anywhere

•Centralised system with patients having personalised dashboard for self monitoring as well as for surveillance by the doctors.

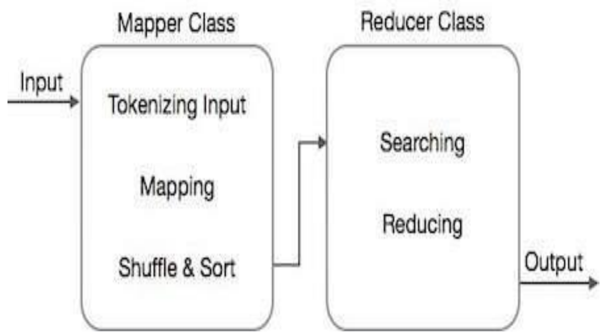
•Socio-demographic factors and locations of patients taped and analysed Thus, this research model can be a great tool in data collection as well as produce real-time data analytics insights

2.EXISTING SYSTEM

In existing system facing problems at large scale performing large-scale computation is difficult. To work with this amount of data requires distributing parts of the problem to multiple machines to handle in parallel. Whenever various machines are used in cooperation with one another, the chance of failures rises [2]. In a single-machine environment, loss is not something that program designers explicitly worry about very often. If the machine has stopped working, then there is no way for the program to recover anyway.

3. LITERATURE SURVEY

The Existing Management System at Murab Hospital, Nigeria. The information flow used is a one directional system where the receptionist refers patient to doctors, doctors referring patients to the pharmacist either in or out patients and the same way out. The system that is currently being used in the hospital is entirely manual [3]. When a patient requests drugs from the employees, all the information is recorded manually from the drug dispenser (Pharmacist). Similarly when the distributor delivers drugs all the information from the dispenser to the account on drugs is recorded manually. The following are the drawbacks of the current system at the hospital:



1. The hospital employees finds it tiresome and time taking when analyzing patient data, drug supplier and staff Payment receipts and voucher cards this leads to delay in medical reports.
2. The hospital Administration currently uses health record files for storing patients and drug supplier's information. This system of data storage is susceptible to security problems such as illegal changes and update of records.
3. The Staff usually waste a lot of time in retrieving data.
4. The paper work reduces the efficiency of the System.

6. SYSTEM ARCHITECTURE:

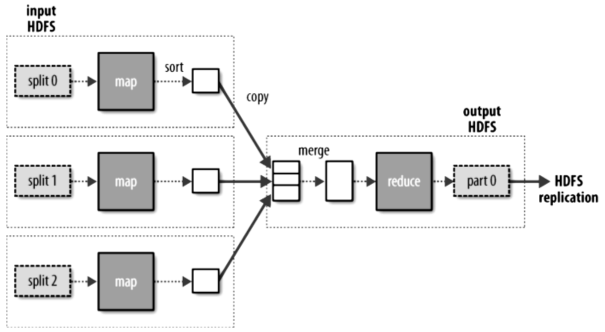
The systems architect establishes the basic structure of the system, addressing the essential core design features and elements that provide the framework for all that follows, and difficult to change later[11]. The systems architect develops the architects view of the users' vision for what the system needs to be and do, and the goes along which it must be able to evolve, and strives to maintain the unity of that vision as it evolves during detailed design and implementation.

4. PROPOSED SYSTEM

The proposed system addresses the problem of data quality in electronic patient records using a computerized patient records report system as an example. Physicians extracted five parameters from a traditional free text report and encoded these parameters thus producing a computer processable report [8]. The proposed system is divided into Receptionist's module, Doctor's module and Pharmacist's module.

5. MAPREDUCE ALGORITHM

Map Reduce works by dividing the processing into two phases: the map phase and the reduce phase. Every phase has key-value pairs as input and output, the types of which may be taken by the programmer. The programmer also defines two functions: the map function and the reduce function



7.B IG DATA IN PROPOSED SOLUTION

The volume of worldwide healthcare data in 2012 was 500 petabytes. That is estimated to grow in 2020 to 25,000 petabytes - a fiftyfold increase from 2012 to 2020[11]. Thus, Big Data in healthcare domain has great potential to help derive meaningful insights after analysis. Big Data in Health care can drive clinical decision support, disease surveillance, epidemic control and population health management[12]. The key function of our application being data collection, data collected can be dealt wih big data technologies. In our proposed system, the patient's input and discharge sheets form the big data. Apache Hadoop provides a technology to process these larger volumes of data and also keep the data on the original data clusters. Per terabyte of storage in Hadoop is, on average, 10 times cheaper than a traditional relational data warehouse system[13]. The Hadoop Distributed File System (HDFS) stores data across multiple data nodes in a simple hierarchical form of directories of files. Conventionally, data is stored in 64MB chunks (files) in the data nodes with a high degree of compression. Hadoop uses MapReduce on the databeing processed.

7.1 Analytical Methods

The goal of medical image analytics is to improve the interpretability of depicted contents . Many methods and frameworks have been developed for medical image processing. However, these methods are not necessarily applicable for big data applications.

One of the frameworks developed for analyzing and transformation of very large datasets is Hadoop that employs MapReduce . MapReduce is a programming paradigm that provides scalability across many servers in a Hadoop cluster with a broad variety of real-world applications . However, it does not perform well with input-output intensive tasks. MapReduce framework has been used to increase the speed of three large-scale medical image processing use-cases, (i) finding optimal parameters for lung texture classification by employing a well-known machine learning method, support vector machines (SVM), (ii) content-based medical image indexing, and (iii) wavelet analysis for solid texture classification. In this framework, a cluster of heterogeneous computing nodes with a maximum of 42 concurrent map tasks was set up and the speedup around 100 was achieved. In other words, total execution time for finding optimal SVM parameters was reduced from about 1000 h to around 10 h. Designing a fast method is crucial in some applications such as trauma assessment in critical care where the end goal is to utilize such imaging techniques and their analysis within what is considered as a golden-hour of care. Therefore, execution time or real-time feasibility of developed methods is of importance. Accuracy is another factor that should be considered in designing an analytical method. Finding dependencies among different types of data could help improve the accuracy. For instance, a hybrid machine learning method has been developed in that classifies schizophrenia patients and healthy controls using fMRI images and single nucleotide polymorphism (SNP) data. The authors reported an accuracy of 87% classification, which would not have been as high if they had used just fMRI images or SNP alone. del Toro and Muller have compared some organ segmentation methods when data is considered as big data. They have proposed a method that incorporates both local contrast of the image and atlas probabilistic information. An average of 33% improvement has been achieved compared to using only atlas information. Tsymbal et al. have designed a clinical decision support system that exploits discriminative distance learning with significantly lower computational complexity compared to classical alternatives and hence this system is more scalable to retrieval. A computer-aided decision support system was developed by Chen et al. that could assist physicians to provide accurate treatment planning for patients suffering from traumatic brain injury (TBI). In this method, patient's demographic information, medical records, and features extracted from CT scans were combined to predict the level of intracranial pressure (ICP). The accuracy, sensitivity, and specificity were reported to be around 70.3%, 65.2%, and 73.7%, respectively. molecular imaging and its impact on cancer detection and cancer drug improvement are discussed.

7.2 Data Storage and Retrieval

With large volumes of streaming data and other patient information that can be gathered from clinical settings, sophisticated storage mechanisms of such data are imperative. Since storing and retrieving can be computational and time expensive, it is key to have a storage infrastructure that facilitates rapid data pull and commits based on analytic demands.

With its capability to store and compute large volumes of data, usage of systems such as Hadoop, MapReduce, and MongoDB is becoming much more common with the healthcare research communities. MongoDB is a free cross-platform document-oriented database which eschews traditional table-based relational database. Typically each health system has its own custom relational database schemas and data models which inhibit interoperability of healthcare data for multi-institutional data sharing or research studies. Furthermore, given the nature of traditional databases integrating data of different types such as streaming waveforms and static EHR data is not feasible. This is where MongoDB and other document-based databases can provide high performance, high availability, and easy scalability for the healthcare data needs. Apache Hadoop is an open source framework that allows for the distributed processing of large datasets across clusters of computers using simple programming models. It is a highly scalable platform which provides a variety of computing modules such as MapReduce and Spark. For performing analytics on continuous telemetry waveforms, a module like Spark is especially useful since it provides capabilities to ingest and compute on streaming data along with machine learning and graphing tools. Such technologies allow researchers to utilize data for both real-time as well as retrospective analysis, with the end goal to translate scientific discovery into applications for clinical settings in an effective manner.

7.3 Data Aggregation

Integration of disparate sources of data, developing consistency within the data, standardization of data from similar sources, and improving the confidence in the data especially towards utilizing automated analytics are among challenges facing data aggregation in healthcare systems . Medical data can be complex in nature as well as being interconnected and interdependent; hence simplification of this complexity is important. Medical data is also subject to the highest level of scrutiny for privacy and provenance from governing bodies, therefore developing secure storage, access, and use of the data is very important.

Analysis of continuous data heavily utilizes the information in time domain. However, static data does not always provide true time context and, hence, when combining the waveform data with static electronic health record data, the temporal nature of the time context during integration can also add significantly to the challenges. There are considerable efforts in compiling waveforms and other associated electronic medical information into one cohesive database that are made publicly available for researchers worldwide. For example, MIMIC II and some other datasets included in Physionet provide waveforms and other clinical data from a wide variety of actual patient cohorts.

7.4. Signal Analytics Using Big Data

Research in signal processing for developing big data based clinical decision support systems (CDSSs) is getting more prevalent. In fact organizations such as the Institution of Medicine have long advocated use of health information technology including CDSS to improve care quality. CDSSs provide medical practitioners with knowledge and patient-specific information, intelligently filtered and presented at appropriate times, to improve the delivery of care.

A vast amount of data in short periods of time is produced in intensive care units (ICU) where a large volume of physiological data is acquired from each patient. Hence, the potential for developing CDSS in an ICU environment has been recognized by many researchers. A scalable infrastructure for developing a patient care management system has been proposed which combines static data and stream data monitored from critically ill patients in the ICU for data mining and alerting medical staff of critical events in real time . Similarly, Bressan et al. developed an architecture specialized for a neonatal ICU which utilized streaming data from infusion pumps, EEG monitors, cerebral oxygenation monitors, and so forth to provide clinical decision support . A clinical trial is currently underway which extracts biomarkers through signal processing from heart and respiratory waveforms in real time to test whether maintaining stable heart rate and respiratory rate variability throughout the spontaneous breathing trials, administered to patients before extubation, may predict subsequent successful extubation. An animal study shows how acquisition of noninvasive continuous data such as tissue oxygenation, fluid content, and blood flow can be used as indicators of soft tissue healing in wound care . Electrocardiogrpah parameters from telemetry along with demographic information including medical history, ejection fraction, laboratory values, and medications have been used to develop an inhospital early detection system for cardiac arrest .

A study presented by Lee and Mark uses the MIMIC II database to prompt therapeutic intervention to hypotensive episodes using cardiac and blood pressure time series data . Another study shows the use of physiological waveform data along with clinical data from the MIMIC II database for finding similarities among patients within the selected cohorts . This similarity can potentially help care givers in the decision making process while utilizing outcomes and treatments knowledge gathered from similar disease cases from the past. A combination of multiple waveform information available in the MIMIC II database is utilized to develop early detection of cardiovascular instability in patients. Many types of physiological data captured in the operative and preoperative care settings and how analytics can consume these data to help continuously monitor the status of the patients during, before and after surgery, are described. The potential of developing data fusion based machine learning models which utilizes biomarkers from breathomics (metabolomics study of exhaled air) as a diagnostic tool is demonstrated.

Research in neurology has shown interest in electrophysiologic monitoring of patients to not only examine complex diseases under a new light but also develop next generation diagnostics and therapeutic devices. An article focusing on neurocritical care explores the different physiological monitoring systems specifically developed for the care of patients with disorders who require neurocritical care . The authors of this article do not make specific recommendations about treatment, imaging, and intraoperative monitoring; instead they examine the potentials and implications of neuromonitoring with differeing quality of data and also provide guidance on developing research and application in this area. The development of multimodal monitoring for traumatic brain injury patients and individually tailored, patient specific care are examined in . Zanatta et al. have investigated whether multimodal brain monitoring performed with TCD, EEG, and SEPs reduces the incidence of major neurologic complications in patients who underwent cardiac surgery. The authors evaluated whether the use of multimodal brain monitoring shortened the duration of mechanical ventilation required by patients as well as ICU and healthcare stays. The concepts of multimodal monitoring for secondary brain injury in neurocritical care as well as outline initial and future approaches using informatics tools for understanding and applying such data towards clinical care are described .

As complex physiological monitoring devices are getting smaller, cheaper, and more portable, personal monitoring devices are being used outside of clinical environments by both patients and enthusiasts alike. However, similar to clinical applications, combining information simultaneously collected from multiple portable devices can become challenging. Pantelopoulos and Bourbakis discussed the research and development of wearable biosensor systems and identified the advantages and shortcomings in this area of study. Similarly, portable and connected electrocardiogram, blood pressure and body weight devices are used to set up a network based study of telemedicine. The variety of fixed as well as mobile sensors available for data mining in the healthcare sector and how such data can be leveraged for developing patient care technologies are surveyed .

8. Big Data Applications in Genomics

The advent of high-throughput sequencing methods has enabled researchers to study genetic markers over a wide range of population improve efficiency by more than five orders of magnitude since sequencing of the human genome was completed, and associate genetic causes of the phenotype in disease states. Genome-wide analysis utilizing microarrays has been successful in analyzing traits across a population and contributed successfully in treatments of complex diseases such as Crohn's disease and age-related muscular degeneration.

Analytics of high-throughput sequencing techniques in genomics is an inherently big data problem as the human genome consists of 30,000 to 35,000 genes. Initiatives are currently being pursued over the timescale of years to integrate clinical data from the genomic level to the physiological level of a human being These initiatives will help in delivering personalized care to each patient. Delivering recommendations in a clinical setting requires fast analysis of genome-scale big data in a reliable manner. This field is still in a nascent stage with applications in specific focus areas, such as cancer, because of cost, time, and labor intensive nature of analyzing this big data problem.

Big data applications in genomics cover a wide variety of topics. Here we focus on pathway analysis, in which functional effects of genes differentially expressed in an experiment or gene set of particular interest are analyzed, and the reconstruction of networks, where the signals measured using high-throughput techniques are analyzed to reconstruct underlying regulatory networks. These networks influence numerous cellular processes which affect the physiological state of a human being

CONCLUSION:

In this paper, we addressed Automated Healthcare Management System is a project implemented with Apache Hive, an abstraction of Map reduce. The data what you are going to analyze is a Semi-structured data. Computerized HMS has been developed. The system solved the problems associated with the existing manual system. Security is also enhanced since access to the system requires authentication. However, the system does not alert the pharmacy of the expiry date of drugs[12]. Also, departments such as security and assets are not included in the design. Therefore, implementing an HMS that can alert the pharmacist of the expiry date of drugs at a given time and manage all departments in the hospital will be an attractive research in future.

9. FUTURE ENHANCEMENTS:

With the development of Apache Hadoop is a next-generation enterprise data architecture is emerging that connects the systems powering business transactions and business intelligence. Hadoop is uniquely implementing storing, aggregating, and refining multi-structured data sources into formats that fuel new business insights. Apache Hadoop is rapidly becoming the important platform for processing Big Data.

Hadoop started from a relatively humble beginning as a point solution for small search systems. Its growth into an important technology to the broader enterprise community dates back to Yahoo's 2006 decision to evolve Hadoop into a system for solving its internet scale big data problems. Eric will discuss the current state of Hadoop and what is coming from a development standpoint as Hadoop evolves to meet more workloads. India takes the second place in the world in its population .Increasing population in India over-burdens the health care structure in the country. We see that there is huge incoming data in the health domain. Thus, the proposed model is the solution for efficient data collection, healthcare delivery integrated with analytics. Thus, this system can automate thehealth care services for patients as well as doctors.

REFERENCES:

[1] Bittencourt, L.F. and Madeira, E.R.M. "A Performance-Oriented Adaptive Scheduler for Dependent Tasks on Grids," Concurrency and Computation: Practice and Experience.

[2] Caron, E. Chis, A. Desprez, F. and Su, A. "Design of Plug-in Schedulers for a GRIDRPC Environment," Future Generation Computer Systems, vol. 24, no. 1, pp. 46-57.

[3] Dinda, P.A. and O'Hallaron, D.R. "Host Load Prediction Using Linear Models," Cluster Computing, vol. 3, no. 4, pp. 265-280.

Dinda, P.A. and O'Hallaron, D.R. "Host Load Prediction Using Linear Models," Cluster Computing, vol. 3, no. 4, pp. 265-280.
[4] Dinda, P.A. "Design, Implementation, and Performance of an Extensible Toolkit for Resource Prediction in Distributed Systems," IEEE Trans. Parallel and Distributed Systems, vol. 17, no. 2,b pp. 160-173.

[5] Eddy Caron , Andreea Chis , Frederic Desprez , Alan Su (November 2011)"Plug-in Scheduler Design for a Distributed Grid Environment".

[6] Liang Hu, Xi-Long Che, (2012) "Online System for Grid Resource Monitoring and Machine Learning-Based Prediction" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 23.

[7] Massie, M.L. Chun, B.N. and Culler, D.E. "The Ganglia Distributed Monitoring System: Design, Implementation, and Experience," Parallel Computing, vol. 30, no. 7, pp. 817-840.

[8] Peter Dinda, A. and David R. O'Hallaron (July 2012) " AN Extensible Toolkit for Resource Prediction in Distributed Systems" School of Computer Science Carnegie Mellon University Pittsburgh, PA ,15213.

[9] Sam Verboven, Peter Hellinckx, Frans Arickx and Jan Broeckhove (2011) "Runtime Prediction based Grid Scheduling of Parameter Sweep Jobs" University of Antwerp Antwerp, Belgium.

[10] Sodan, A.C. Gupta, G. Han, L. Liu, L. and Lafreniere, B. "Time and Space Adaptation for Computational Grids with the ATOPGrid Middleware," Future Generation Computer Systems, vol. 24, no. 6, pp. 561-581.